

## Chapter Three: Adaptive Preferences

The welfarist's dilemma, once again, runs like this: for a welfarist approach to succeed, the basic minimum must be tied closely to the achievement of one's conception of the good. Otherwise, Nussbaum's argument for a capabilities approach sticks: a welfarist approach to the basic minimum *would* insist on a moral reason to force individuals to live lives they do not value. WBM successfully avoids this concern: the basic minimum is not construed as the achievement of particular states or functionings in abstraction from conceptions of the good. The basic minimum is the achievement of a *valued* project.

So far so good. But the second horn of the dilemma looms. To be a valued project, as noted in the last chapter, crucially depends on the attitudes one takes toward this project and its context. But such attitudes are malleable, often in insidious ways. The fact that one values  $p$  can be shaped by poor circumstances and a lack of opportunity. If so, one should think that the mere fact that A values a particular project is not enough for A to maintain the basic minimum. A's valuing might be the product of adaptive preferences.

The phenomenon of adaptive preferences clearly poses *some* problem for WBM. But the problem it poses is ill-understood and under-theorized. In this chapter, I distinguish two categories of adaptive preferences: *shallow* adaptive preferences, and *deep* adaptive preferences. Shallow adaptive preferences and deep adaptive preferences, I shall argue, pose very different problems for a welfarist approach to the basic minimum. My approach can successfully avoid the problem posed by shallow adaptive preferences. And while it cannot solve the problem of deep adaptation, this should not be held against my view. I argue that deep adaptation is no problem *in itself*.

This chapter is organized as follows. §3.1 discusses the mechanism by which Nussbaum's capabilities approach addresses cases of adaptation. In §3.2, I address the problem of adaptive preferences in more detail. I argue that this problem, far from being a reason to reject welfarism, is actually a problem *internal* to welfare theory. No theory of well-being can possibly be adequate if this theory holds that adaptive preferences improve lives. Indeed, I argue that the problem posed by adaptive preferences (or "shallow" adaptive preferences, at any rate) is that such preferences do not properly characterize a person's genuine conception of the good. In §§3.3-3.4, I offer an account of conceptions of the good that avoids the problem of adaptation. In §3.5 I discuss the problem of deep adaptation; I argue that this problem simply reduces to a problem I call "preference for the worse". As such, it

permits of easy solution (if, indeed, a solution is called for at all).

One note before I begin. Generally speaking the phenomenon to which I refer is called “adaptive preferences”. I follow the literature on this point. But I treat the term “preference” as a catch-all for pro-attitudes that figure into a person’s wider conception of the good (including pro-attitudes that help to determine whether particular projects are valued, e.g., an assessment that *p contributes to a life worth living*, etc.). This makes no difference to the argument of this chapter, however; all such attitudes are subject to adaptation, and my solution applies *mutatis mutandis*.

### 3.1. Nussbaum on Adaptation

Preferences are malleable. But this malleability can cause serious problems for any view that takes the fulfillment of a person’s preferences as an important aspect of the basic minimum. The most infamous examples are offered by Amartya Sen: “The hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the over-exhausted coolie may all take pleasures in small mercies, and manage to suppress intense suffering for the necessity of continuing survival, but it would be ethically deeply mistaken to attach a correspondingly small value to the loss of their well-being because of this survival strategy.”<sup>1</sup> As Sen notes, the “landless laborer” might adapt to his poor circumstances: because a life of landless labor is the only life to which he has access, he comes to adjust his desires and expectations to his straightened circumstances. Hence, one might think, to index the achievement of the basic minimum to these preferences is “ethically deeply mistaken”.

In justifying the rejection of a welfarist approach to the basic minimum, Nussbaum notes the following ways in which preferences can be adaptive: “The normative approach based on human functioning and capability... rejected utilitarian preference-based approaches as a basis for fundamental political principles precisely because they were unable to conduct a critical scrutiny of preference and desire that would reveal the many ways in which habit, fear, low expectations, and unjust background conditions deform people’s choices and even their wishes for their own lives.”<sup>2</sup> Elsewhere, Nussbaum writes that the problem of adaptation boils down to the fact that, in cases of adaptation, “an individual’s preferences are shaped to accord with the (frequently narrow) set of opportunities she actually has.”<sup>3</sup>

---

<sup>1</sup>Sen (1987), 45-6, my emphasis.

<sup>2</sup>Nussbaum (2000), 114.

<sup>3</sup>Nussbaum, “American Women: Preferences, Feminism, Democracy” in Nussbaum

For Nussbaum, adaptive preferences are shaped by the *particular facts* of an individual's lived life, especially that person's opportunities (or lack thereof). The mechanism by which one's preferences adapt to the "set of opportunities" one actually has can vary. Preferential adaptation can occur as a result of a number of psychological mechanisms: a "survival strategy", as noted by Sen; a lack of "self-worth" as discussed by Jennifer Hawkins;<sup>4</sup> habit, indoctrination, strategic preference engineering, and perhaps many others.

Nussbaum holds that the capabilities approach, unlike "subjective welfarism", can avoid the problem of adaptive preferences. First, for Nussbaum, the basic minimum is not dependent on the fulfillment of preferences. Rather, the basic minimum is identified as the maintenance of the ten central capabilities. An individual must possess these to maintain the basic minimum, and will maintain the basic minimum once these central capabilities are possessed, no matter the content of his or her preferences or pro-attitudes. Hence when it comes to the basic minimum, at least, moral reasons are not sensitive to the content of any individual's preferences, unlike under WBM. Nussbaum writes: "[T]he capabilities account deals well with the problems that plagued the preference-based approach... A habituated preference not to have any one of the items on the list (political liberties, literacy, equal political rights, or whatever) will not count in the social choice function."<sup>5</sup> Hence, for Nussbaum, the problem of adaptive preferences is avoided by refusing to index the achievement of the basic minimum to preferences.

However, this response by itself is not satisfying. To see this, consider the central problem for WBM. The only reason WBM might fail in light of the phenomenon of adaptive preferences is that WBM holds that people who live shabby lives as a result of adaptive preferences will, on occasion, obtain the basic minimum. In saying this, WBM is "ethically deeply mistaken". But there is no guarantee that the same will not hold of the capabilities approach. Simply declaring that the ten basic capabilities form the content of the basic minimum still leaves open the possibility that people will maintain, and act on, adaptive preferences, i.e., that people (like the "overexhausted coolie") will live horribly shabby lives despite the fact that they maintain the basic minimum. Without some further mechanism to *overcome* the psychological entrenchment of adaptive preferences, the capabilities approach is impotent to respond to the problem that plagues WBM. Any person who lives a

---

(1999), 151.

<sup>4</sup>See Hawkins (2008).

<sup>5</sup>Nussbaum (2000), 149.

shabby life as a result of adaptive preferences might nevertheless also do so *even if* they secure the ten basic capabilities. If so, people who live shabby lives as a result of adaptive preferences will, at least on occasion, maintain the basic minimum. If this is a problem for welfarism, it is no less a problem for the capabilities approach.

Nussbaum is aware that leaving the matter here is unsatisfactory. Remember that, for Nussbaum, adaptive preferences are a *result* of narrow opportunities or capabilities. If so, there is no reason to fear a refusal to develop valuable functionings once the central capabilities are granted. For Nussbaum, the capabilities *themselves* are mechanism enough to overcome problems of adaptation insofar as adaptation is specifically a result of a lack of capabilities. After all, the overexhausted coolie develops his preferences in light of the fact that “truly human functioning” is not available. *Were* it available, Nussbaum surmises, the overexhausted coolie would not come to prefer the life of overexhausted coolieism.<sup>6</sup> In this way, the capabilities approach not only insists that all persons should be granted the ten basic capabilities, but the capabilities themselves work to combat the existence and development of adaptive preferences. For Nussbaum, the various psychological mechanisms that result in adaptive preferences share a common cause. Adaptive preferences are, at base, the result of facts of unavailability; facts that, if alleviated, would cause us to revert to non-adaptive preferences. In this way, and only in this way, does the capabilities approach alleviate the problem of adaptive preferences.

### *3.2. Adaptive Preferences and Well-Being*

If WBM allows adaptive preferences to determine the content of anyone’s basic minimum, it is in serious trouble. Many of the most important cases that motivate a moral commitment to a basic minimum involve people who live shabby lives, cannot get sufficient nourishment, are imprisoned, or desperately sick. However, given that such people can *adapt* to their straightened circumstances and hence come to value whatever projects they may achieve in such circumstances, it would appear that any welfarist basic minimum is in tension with a major motivation for a basic minimum: to assist those who live shabby lives, cannot get sufficient nourishment, are imprisoned, or desperately sick. Hence because someone might achieve a valued project that is valued only as a result of problematic adaptation (say, for instance, the valued project of being “an overexhausted coolie”), and because WBM

---

<sup>6</sup>See Nussbaum (2000), 155.

only entails a moral reason to promote the achievement of valued projects, there is no mechanism by which WBM might avoid the problem of adaptive preferences. Hence WBM cannot be trusted to properly account for the basic minimum. In this way, the problem of adaptive preferences can be seen as another way of stating a (by now) familiar objection: because many people can adapt to ridiculously poor lives and hence come to adaptively see them as worth living, WBM sets the minimum threshold too low.

But though this may be a problem for WBM, the root of the problem is not simply that WBM is a welfarist approach. We describe the problem of adaptive preferences inaccurately if we say that it is a problem for welfarism generally. Rather, the problem of adaptive preferences is itself *internal* to an inquiry into the nature of human well-being. The fulfillment of adaptive preferences, so I shall argue, does not improve lives. The problem for WBM is its implicit commitment to the claim that the satisfaction of adaptive preferences *is a prudential benefit*.<sup>7</sup> But this is a false claim *about human well-being*.

By way of an example, consider the following case:

*Erin*: Erin, at  $t_1$ , intrinsically preferred a life of excellent achievement in dance to a life of excellent achievement in painting and dedicated years to the achievement of her goal. At  $t_1$ , Erin regarded the movement of the human form as a much more meaningful expression of her artistic temperament than mere paint on canvas. However, as a result of years of rejection and failure, she came to a decision. To avoid continued frustration and regret, she would attempt to alter her preferences away from dancing and toward painting. Suppose she is successful: Erin, at  $t_2$ , prefers the life of painting to the life of dance.

Erin's case is an example of adaptive preferences. Erin, merely because it is available, comes to prefer a life of painting to a life of dance. Hence any theory that draws an important connection between Erin's adaptive preferences and her good must claim that a life of painting is better for Erin at  $t_2$  than a life of dance. But this is counterintuitive. Her preference

---

<sup>7</sup>Indeed, Nussbaum holds something similar. Nussbaum writes: "One might hold that the preference-based view is perfectly all right as a basis for quality of life comparisons, while doubting that it could be an adequate basis for the selection of basic political principles. . . This would, I think, be an implausible position to hold, once one confronts the specific defects of preference-based views: the reasons for thinking them bad bases for political principles are also reasons why they do not do a good job of quality of life measurement," (Nussbaum (2000), 116).

is an example only of a successful attempt to avoid the predictable and painful frustration that comes along with a preference for that which is simply unavailable. Of course, we are perfectly licensed to say, at  $t_2$ , that a life of painting would improve Erin's life over the lack of either activity. Furthermore, it seems right to say, of Erin, that *given* her lack of talent, it is prudentially *rational* to switch her preferences in this way. But insofar as her preferences are merely a result of strategic engineering, of intentional adaptation to her circumstances, it seems wrong to take them at face value. Contrary to her preferences post-engineering, *were* she able to be a dancer at  $t_2$ , it seems quite right to say that the life of a dancer would be better for Erin at  $t_2$  than the life of a painter. Thus it seems wrong to say that Erin's  $t_2$  preferences should be authoritative over that which is good for her at  $t_2$ .<sup>8</sup> Any theory that admits that the fulfillment of adaptive or distorted preferences improve lives is thus mistaken. Hence WBM has two strikes against it. Not only is it "ethically deeply mistaken" as a theory of the basic minimum, it also seems to rely on a false theory of welfare: that a person's valuing attitudes (some of which will be adaptive) are a guide to the most central elements of a person's life quality.

One might be tempted to claim that this feature of adaptive preferences is good news for a welfarist approach to the basic minimum, if not for WBM: if adaptive preferences are bad guides to an individual's good, then it must be the case that the *proper* account of that which is in an individual's interest might be able to construct an approach to the basic minimum that avoids the problem at hand. But this possibility is complicated by the first horn of the dilemma I posed in Chapter One: only approaches to a welfarist basic minimum that take an individual's conception of the good (and hence potentially adaptive preferences) seriously in setting the basic minimum can avoid the unacceptable conclusion that people should be forced to live in ways they do not value. Hence if the problem of adaptive preferences cannot be solved by WBM, it seems unlikely than any welfarist approach to the basic minimum will succeed.

To begin my defense of WBM, I note first that though many agree that adaptive preferences cause problems for particular theories of welfare, the

---

<sup>8</sup>That Erin's preference is not authoritative is not evidence that achievement in painting is not good for her. Indeed, Erin may have some *other* preference, viz., the preference to paint rather than *nothing*, which *is* authoritative. Furthermore, one might also claim that achievement in painting is good for Erin no matter what she prefers. To say that a particular preference is not authoritative is simply to say that this preference is not a guide to an individual's good. It does not entail that the preferred *object* is of no welfare value.

precise nature of the problem appears ill-understood.<sup>9</sup> Hence to address the problem of adaptive preferences, I must first investigate *what*, precisely, this problem is. What *about* adaptive preferences renders them irrelevant to welfare?<sup>10</sup> This question will take up the next three sections.

### 3.2.1. Preference for the Worse

First, one might be tempted to claim that adaptive preferences are welfare-irrelevant because they cause a person to *prefer the worse*; that is, to prefer those things that are worse as a matter of *objective* prudential value. Sen's landless laborer, quite plausibly, prefers the worse insofar as he prefers landless labor to a life of greater achievements, education, relationships, etc. These things are objectively better than landless labor. Hence the landless laborer's preferences are welfare-irrelevant, mistaken.

Preference for the worse is obviously an important challenge to views that posit subjectively defined welfare goods. As such, it be dealt with in due time. However, the problem of adaptive preferences should be kept distinct from the problem of preference for the worse. A preference can be adaptive without being an instance of preference for the worse. Erin is a prime example. It strikes me as clearly implausible to say that Erin's post-adaptation preferences are for the worse. Erin's preferences are surely adaptive; indeed, they are consciously, explicitly adaptive. If social policy is interested in improving Erin's life, it will not treat her preference for painting as authoritative. But Erin's preferences are not an example of preferring the worse to the better.

Divorcing the problem of adaptive preferences from the problem of preference for the worse sheds some important light on previous attempts to solve the problem of adaptive preferences. One popular solution is to declare that some things are intrinsically bad and some things are intrinsically good for people independently of anyone's preferences for them. One might, for instance, reject subjectivism about well-being on the whole, or adopt a form of Weak Subjectivism, which has the power to declare that some

---

<sup>9</sup>For vastly different approaches to understanding the nature and problem of adaptive preferences, see Elster (1982) and (1983), esp. ch. 3; Nussbaum, "Women and Cultural Universals" in (1999) and (2000), esp. ch. 2; Sumner (1996), esp. ch. 6; Sen (1992), 6-7; Rickard (1995); Qizilbash (2006).

<sup>10</sup>I use the term "welfare-relevant" for any preference or element of an individual's conception of the good that can sensibly play a role in determining that which is good for a person; welfare-relevant preferences need not determine the all-things-considered betterness of a preferred object; they are simply one factor among others in determining a person's well-being.

things people prefer are intrinsically worthless. On this view, any preference for, say, overexhausted coolieism is not relevant to welfare; overexhausted coolieism is not a prudential benefit no matter how much it is preferred. However, this solution is insufficient to solve the problem at hand.<sup>11</sup> It conflates the problem of adaptive preferences with the problem of preference for the worse. This conflation leads to a failure to address cases of adaptive preferences that do not *also* involve cases of preference for the worse. Erin's is one such case. It is wildly implausible, it seems to me, to declare that a life of excellent achievement in painting is objectively worse than a life of excellent achievement in dancing. But unless the view in question is willing to take the implausible step of declaring that neither possibility is better for Erin, it cannot simply reject *tout court* the importance of Erin's preferences to her well-being. But any view that takes the plausible step of allowing the occasional authority of preference remains vulnerable to the problem of adaptive preferences whether or not it solves the problem of preference for the worse. Though preference for the worse is an important problem in its own right, solving this problem does not entail solving the problem of adaptive preferences.

### 3.2.2. *Adaptation and Autonomy*

The problem of adaptive preferences goes beyond preference for the worse. But what, then, is the problem? Some have claimed that the distinctive problem illustrated by the phenomenon of adaptive preferences is that some preferences are *non-autonomous*, and hence any theory of well-being that holds that such preferences are welfare-relevant is out of tune with a person's autonomy or autonomous nature. As Sumner writes:

Why are we reluctant to take at face value the life satisfaction reported by 'the hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the over-exhausted coolie'? Presumably because we suspect that the standards which their self-assessments reflect have been artificially lowered or distorted by processes of indoctrination or exploitation. In that case, the obvious remedy is to correct for

---

<sup>11</sup>Hawkins proposes a solution that accepts that certain lives can be *bad* for persons regardless of their own preferences, but cannot be *good* for persons without the authority of their preferences. Her proposal accepts a form of Weak Subjectivism. See Hawkins (2008), 167-8. Others have adopted objective theories of well-being to solve the problem of adaptive preferences. See, for instance, Qizilbash, *op. cit.* and Rickard, *op. cit.* Both strategies fail.

the conditions under which their expectations about themselves came to be formed. The problem is not that their values are objectively mistaken but that they have never had the opportunity to form their own values at all. They do not lack enlightenment, or insight into the Platonic form of the good; they lack autonomy.<sup>12</sup>

Sumner speaks of “happiness” or “life satisfaction”. But leaving this aside for the moment, a “Sumnerian” approach to adaptive preferences might hold that the problem to be illustrated by adaptive preferences is that adaptive preferences are not properly *autonomous*.<sup>13</sup> For Sumner, problematic forms of adaptation are a result of non-autonomous mechanisms, both social and psychological. On this view, one solves the problem of adaptive preferences by claiming that preferences formed by non-autonomous processes are welfare-irrelevant, preferences that are—or would be—formed by autonomous processes are accepted as welfare-relevant.

Missing from this proposal is a theory of autonomous preference formation. Sumner does not offer such a theory, and does little to clarify what it means for a preference to be formed by a non-autonomous process. However, he does offer a list of preference (or “happiness”) formation processes that he regards as non-autonomous

Self-assessments of happiness or life satisfaction are suspect (as measures of well-being) when there is good reason to suspect that they have been influenced by autonomy-subverting mechanisms of social conditioning, such as indoctrination, programming, brainwashing, role scripting, and the like. [...] [T]he best strategy here is to treat subjects’ reports of their level of life satisfaction as defeasible—that is, as authoritative unless there is evidence that they are non-autonomous.<sup>14</sup>

Such socialization processes can surely give rise to adaptive preferences. One might imagine, for instance, that the landless laborer’s life of poverty is an importantly non-autonomous preference formation process. But the problem of adaptive preferences cannot be solved by claiming that preferences or reports of satisfaction or happiness are non-autonomous. *Some* psychological mechanisms by which adaptive preferences are formed will be perfectly compatible with a person’s autonomy. For instance, it is hard to see Erin’s

---

<sup>12</sup>Sumner (1996), 166.

<sup>13</sup>This view is also endorsed by Elster in Elster (1982).

<sup>14</sup>Sumner (1996), 171.

preference revision as non-autonomous.<sup>15</sup> Erin prefers the life of a dancer for its own sake, but finds that she has no talent for it. She then, quite rationally, decides to stop pining away for the life she longed for, and strategically attempts to revise her preferences toward the available: the life of a painter. Indeed, far from being an instance of pernicious socialization, embarking on a strategy of preference revision of this kind is straightforwardly rational. Why continue to pine away for something when it is clearly beyond your grasp? Preference switching in this case will lead to far fewer cases of frustration and regret which is surely, all things considered, something to be praised and encouraged.<sup>16</sup>

One might reply that if they are the result of a rational or autonomous of preference formation, we have no reason to believe that, e.g., Erin's  $t_2$  preferences are not a guide to her good. This suggestion is mistaken. Merely because it is *rational* to at least attempt to switch her preferences, or that in so doing she would be acting autonomously, does not mean that the life of a dancer wouldn't be better for Erin. If she *could* achieve it, we are tempted to say, it would certainly be better for her. Erin's adaptive preferences interrupt the inference from her preference to her good, but not because the process by which the preferences were formed is somehow irrational or non-autonomous.

### 3.2.3. *Adaptation and Proto-autonomy*

Adaptive preferences are welfare-irrelevant not because they are instances of preference for the worse or because they are non-autonomous. But if this is correct, we appear to lack justification for treating Erin's preferences as irrelevant to her good. One might put this point in the following way: once preference for the worse and failures of autonomy have been swept away, why should we believe that, e.g., Erin's welfare at  $t_2$  is determined, not by her preferences at  $t_2$ , but rather by her preferences at  $t_1$ ? On the one hand, it seems right to say that Erin's preferences ought to have some authority over her well-being, especially given that dancing and painting are both worthwhile activities. But on the other hand, we seek to reject the authority of her  $t_2$  preferences given their adaptation. We seem to want Erin's preferences to determine her welfare, and then refuse to grant her preferences such authority. What gives?

To answer this question, we must answer a much more fundamental

---

<sup>15</sup>Compare Rickard (1995), 289-90.

<sup>16</sup>Cf. Rosati (2006), esp. 60-64.

one. Why should we believe that a person's preferences should have *any* welfare-determining authority at all? Surely at least *part* of the motivation for accepting the prudential authority of preferences is a commitment to the claim that what is good for a person should be—perhaps even only to a very minor degree—determined by his or her *conception of the good*, or what this person, in the broadest sense, values.<sup>17</sup> And if we are interested in tying Erin's good to her conception of the good, we must say that preferences have at least some authority over her welfare. But this sheds new light. In Erin's case, it seems right to say that her preference at  $t_2$  for painting over dancing does not really capture her conception of the good: her preferences are merely a result of an unusually successful attempt at strategic preference engineering. If so, *Erin's adaptive preferences are not part of her genuine conception of the good*, and hence does not really capture what we care about when we seek to grant welfare-determining authority to preferences.

This verdict appears to generalize: adaptive preferences show that any old preference need not always express its possessor's conception of the good. This holds just as well for Sen's overexhausted coolie as it does for Erin. Though the overexhausted coolie might prefer coolieism as a survival strategy; overexhausted coolism is not a feature of his true *values*. It is a product, rather, of manipulation by external circumstances, such as a lack of available alternatives. Though it is correct to say that at least some of Erin's preferences shift from  $t_1$  to  $t_2$ , such a shift need not entail that her conception of the good *also* shifts. Rather, they are a mere product of circumstance; they reflect, as it were, a strategic "coping mechanism".

My examples here are stylized, but the point holds. It seems right to say that the overexhausted coolie, the landless laborer, etc., do not *really* value the circumstances they are in or the projects they might engage. Their preferences are a result of a coping mechanism, or simple habit, or a form of indoctrination. Indeed, Nussbaum's own solution to the problem of adaptive preferences relies on this diagnosis. For it to be the case that greater capabilities will yield the reversal of adaptive preferences, Nussbaum must believe that a person's adaptive preferences are not a real part of her conception of the good, but are a mere product of a lack of alternatives. Otherwise there is no reason to believe that additional capabilities will alter the life she leads. To put this another way, in cases of adaptation, *real* preferences are "covered up" by the lack of capabilities, and hence Nussbaum is licensed to

---

<sup>17</sup>Again, this claim is perfectly compatible with subjectivist and objectivist theories of the good; objectivist views can claim, plausibly, that the fulfillment of a person's preferences *occasionally* improves their welfare, perhaps when the preference fulfilled is for a valued project.

expect that *given* further opportunities, people will overcome adaptation, and come to prefer lives that enjoy “truly human functioning”.

If my analysis is correct, there need be no tension between the project of avoiding adaptive preferences and the project of accommodating *The Basic Minimum as Proto-autonomy*. If adaptive preferences do not reflect an individual’s true conception of the good, the project of constructing a basic minimum that does not succumb to the problem of adaptive preferences just *is* the project of accommodating *The Basic Minimum as Proto-autonomy*. Hence, or so it would appear thus far, all that is required to defeat adaptive preferences is to account for a person’s “true” conception of the good—a conception of the good not biased toward the status-quo or distorted by facts of unavailability, habit, strategic engineering, etc. If such an account can be found, WBM is under no threat.

### 3.3. *Conceptions of the Good: A Theory*

Time to take stock. The problem for WBM seemed to be this: according to WBM, one maintains the basic minimum insofar as one maintains a valued project. But one’s “values” can be based on preferences or other pro-attitudes (such as an assessment of a life as worth living, etc.) that are susceptible to adaptation. However, I have so far argued that adaptive preferences do not reflect an individual’s “true values” or true conception of the good. Hence to complete the project of accommodating *The Basic Minimum as Proto-autonomy*, and to avoid the problem of adaptive preferences, we simply need a better account of an individual’s true preferences that avoids the mechanisms of adaptation. If this can be offered—and this is a big “if”—WBM is under no threat.

Call my account of a person’s genuine preferences “preference coheren-tism” (PC). I won’t be able to defend all the nooks and crannies of PC here; I set out only to show how it might solve the problem of adaptation. There are surely powerful objections to my view; I leave these aside. (For instance, I will not defend my precise conception of coherence. Furthermore, some have argued that my account is circular. I argue elsewhere that it is not, but will leave aside this objection here.<sup>18</sup>) PC declares that a person’s genuine conception of the good is constituted by her evaluative beliefs, *rendered coherent and complete*. I present this account in four stages. First,

---

<sup>18</sup>I try to specify a working notion of coherence good enough for our purposes here in Dorsey (2006). For a defense against the charge of circularity, see Dorsey (2010b) and (MS).

I argue that genuine preferences should be understood as evaluative beliefs rather than desires. Second, I more fully describe the “coherence and completeness” constraints. Third, for the purposes of clarification, I compare my account of valuation to other superficially similar accounts, including a coherentist moral epistemology, and an account of prudential valuing offered by James Griffin. Finally, I argue that PC solves the problem of adaptive preferences.

### 3.3.1. *Beliefs not Desires*

The first noteworthy feature of preference coherentism is that it understands preferences or other pro-attitudes in terms of evaluative beliefs rather than desires. According to PC, A prefers  $p$  to  $q$  to the extent that A believes (under the right conditions) that  $p$  is better for A than  $q$ . Understanding preferences in terms of beliefs rather than desires requires defense. Most importantly, the appeal to beliefs rather than desires in defining welfare-relevant preferences can correct the influence of certain adaptation-causing psychological mechanisms. This proposal is intuitive, as noted by Sidgwick: “a prudent man is accustomed to suppress, with more or less success, desires for what he regards as out of his power to attain by voluntary action—as fine weather, perfect health, great wealth or fame, etc.; but any success he may have in diminishing the actual intensity of such desires has no effect in leading him to judge the objects desired less ‘good.’”<sup>19</sup> Here Sidgwick notes that one adaptation-causing psychological mechanism is “prudence”, or rational suppression of a desire for that which is unavailable. (One might imagine that this mechanism is at work in Erin’s case.) But this form of suppression occurs only at the level of desire: according to Sidgwick, this form of adaptation “has no effect in leading him to judge the objects desired less ‘good.’”

Other mechanisms by which adaptive preferences may form are limited to desire. To see this, consider Jennifer Hawkins’ discussion of a character in Alice Walker’s *The Color Purple*:

Celie’s early life is a study in trauma and abuse. From a young age, she is sexually abused by her stepfather (whom she falsely believes to be her biological father). . . Eventually, Celie escapes this tormentor for a new one—the man her stepfather forces her

---

<sup>19</sup>Sidgwick (1907), 110. Sidgwick’s own account, however, fails to properly address this problem: Sidgwick himself insists that a person’s desires are authoritative over his good only among those choices that are “open to him”. See Sidgwick (1907), 111-12.

to marry and who is simply referred to throughout the novel as Mr. —. Mr. — treats Celie like a slave, demanding not only that she care for his unruly children from a previous marriage and maintain a high standard of housekeeping, but that she work in his fields as well. He beats her mercilessly and constantly tells her how stupid and ugly he thinks she is.<sup>20</sup>

Celie comes to accept this treatment as her lot in life; in describing her condition, Hawkins writes as follows:

One particularly striking feature of... Celie [is her] low sense of self-worth. As I suggested before... Celie *feels* worthless. Without a proper sense that she, Celie, matters, her idealized self will be unlikely to give her non-idealized self good advice. Nor will she be able to assess her life fairly in comparison with other lives. Although she may recognize that other lives are happier or more successful, she will not necessarily see them as better for her if she herself feels unworthy of the lives in question.<sup>21</sup>

As Celie seems to clearly illustrate, some people, because of their circumstances, will develop a lack of self-worth and will, as a result, develop preferences that are adaptive or biased toward the status-quo. This might occur not only for people whose lives are objectively bad (such as Celie or the landless laborer), but also in others; one could imagine, for instance, that Erin's preference for painting might arise given a conviction that she is "unworthy" of the life of a dancer. But it is important to note that there are (at least) two ways in which low self-worth could lead to adaptive preferences. First, a person with low self-worth might *believe* that other lives would be substantially better than her current life, but because she has a low sense of self-worth and "feels unworthy of the lives in question" she fails to desire them. Second, a person with low self-worth might simply come to judge that the life in which she is, say, dominated, landless, overexhausted is actually good for her, perhaps because it somehow "fits" her low conception of her own worth.

I will leave discussion of the second possibility until §3.4. But PC's emphasis on beliefs rather than desires helps to address the first permutation of Hawkins' treatment of adaptation: even though Celie might believe herself unworthy of that which she believes is good, she still has a conception of what is good for her—she *believes* that lives other than her own are good

---

<sup>20</sup>Hawkins (2008), 145-6.

<sup>21</sup>Hawkins (2008), 160-1.

for her, or might be better than the one of which she feels herself worthy. Her lack of self-worth causes adaptive preferences only at the level of desire: because she lacks self-esteem, she fails to *want* that which she judges good for her. Though this might not be the only way in which low self-worth can contribute to adaptive preferences, it is certainly *one* way it might do so. Hence understanding genuine preferences in terms of evaluative beliefs rather than desires better explains a person's *true* values. Thus it seems to me there is good reason to understand a person's conception of the good in terms of what this person believes is good for her.

### 3.3.2. Coherence and Completeness

Understanding a person's preferences in terms of evaluative belief rather than desire goes some distance toward solving the problem of adaptive preferences. But it does not go far enough. An account of a person's genuine preferences must be abstracted from the person's actual evaluative beliefs, and must be identified with what a person *would* come to believe under certain counterfactual conditions. There are two reasons for this. First, a person's conception of the good might be incoherent, inconsistent, or self-refuting. Second, without such abstraction we leave open the possibility of adaptation. Certainly psychological mechanisms exist (including, e.g., rational preference revision) that could render not just desiderative preferences, but also one's evaluative beliefs, adaptive to one's—potentially quite shabby—circumstances. Taking evaluative beliefs at face value, then, fails to eliminate all mechanisms by which preferences might be adapted, and hence fails to properly accommodate an interest in proto-autonomy.

A person's conception of the good must be abstracted from her actual evaluative beliefs in two ways. First, genuine preferences must be understood not as that which she happens to believe is good, but as that which she *would* believe good *were* her conception of the good rendered *coherent*. Incoherence involves not only contradictory beliefs, but beliefs that are ill-behaved in various ways, including those that display intransitivity.<sup>22</sup> Furthermore, coherent beliefs support each other and provide explanatory and justificatory connections.<sup>23</sup> Though I take this to be a rather weak requirement, we

---

<sup>22</sup>Though transitivity has been denied as a feature of the good, I accept it as bedrock. For conflicting views, see Rachels (1998); Temkin (1996). Arguments, in my view successful, for transitivity are to be found in Broome (2004), ch. 4; Norcross (2002).

<sup>23</sup>One such explanatory connection is exploited in the last chapter: an assessment of a life as worth living must be explained, at least in part, by a further judgment of the value of a global project *p* if *p* is to be a valued project.

should at least insist that welfare-relevant preferences that purport to reflect a person's genuine conception of the good be supported and warranted by other members of her conception of the good.<sup>24</sup>

Second, conceptions of the good should be construed as arising from a *complete* set of evaluative beliefs. The completeness requirement is really two requirements in one. The first is that welfare-relevant conceptions of the good must yield a *complete ordering* of all possible welfare goods. (If commensurability is limited, A's coherent and complete conception of the good should yield *as complete an ordering as possible*.) By "possible", I mean *metaphysically possible*. Insofar as we believe that a change in A's capacities, circumstances, or environment might be good or bad for A, we shouldn't hold these capabilities, circumstances, or environment fixed when issuing a complete account of that which would be better or worse for A. This is a crucial step in avoiding preference adaptation: if one's conception of the good only rank-orders activities or other welfare goods that are possible only in some more restricted sense of "possible" (such as physically possible, economically possible, etc.), this will open the door to adaptation to that which is possible in this more restricted sense. But insofar as any set of preferences are to be a reliable guide to a person's good (at least some of the time), and insofar as we believe that any metaphysically possible life *l* could be better or worse for a person, that person's conception of the good should evaluate *l*.

However, to guarantee a complete ordering, a conception of the good must have a sufficient basis to determine a complete ordering without gaps. This consideration naturally leads into the second half of the "completeness" requirement. Evaluative beliefs must be tested against a complete set of *value data*. Complete testing closes potential gaps. Though I now have no beliefs that will yield a proper ordering between the life of an Aztec chieftain or a Mayan chieftain, testing my evaluative beliefs against the relevant value data will close this gap (or, if they are genuinely incommensurable, declare them so). Thus the completeness requirement refers both to the mandated complete ordering, and also to the requirement that any genuine conception of the good survive a counterfactual process of complete testing.

What are value data? On my understanding, a value datum consists of two crucial elements. First, the *information* about *what living a given life would be like*. A value datum will require the full confrontation with the consequences, experiences, achievements, etc., of living some life. There are different ways one might construe this element of value data. One might

---

<sup>24</sup>I argue for this in more detail in Dorsey (2006).

conceive of it as a “report”, i.e., some list of facts about the content of a life, or as an “experience”, i.e., the actual *experience* of living a life reported on in the report model.<sup>25</sup> According to David Sobel, the choice is not inconsequential: a mere report, in comparison to the full experience, will fail to accurately convey the bases for a complete evaluation.<sup>26</sup> Sobel’s critique seems correct, and hence I will accept it for my purposes here. Thus value data will require full information about what a given life would be like for the person who lives it, conceived of as the experience of actually living it. The second element of value data is a *judgment* about the quality of that life *given* the experience of it. Thus value data is properly conceived of as a belief in the quality of some particular life on the basis of actually experiencing that life, *and on that basis only*. A completely tested conception of the good, then, will be tested against judgments of the quality of all metaphysically possible lives.

One point remains. The coherence and completeness requirements introduce the possibility of recalcitrant data: a belief about the good that conflicts, or renders incoherent, the set of evaluative beliefs it is used to test. If so, revisions to this set with an eye toward renewed coherence will be required. How are we to go about rendering coherent an incoherent set? Here PC is conservative. When revising in light of recalcitrant value data or as a result of incoherence, revisions are made at the “periphery”, revising a person’s most strongly held preferences only as a last resort. Thus, putting all this together, PC holds that A’s genuine conception of the good is constituted by A’s set of beliefs about that which is good for A, after having been tested against all possible value data, and revised in light of recalcitrant data and other forms of incoherence by a process of “minimal mutilation”.<sup>27</sup>

### 3.3.3. *Some Comparisons*<sup>28</sup>

PC is a theory of a person’s genuine preferences. But before I attempt to show that this account solves problems of adaptation, I should compare my view to some that are similar in certain ways, dissimilar in others. Those uninterested in such comparisons can proceed to the next section.

First, David Brink outlines a coherentist moral *epistemology* identified as follows: “Moral coherentism or a coherence theory of justification in

---

<sup>25</sup>Sobel (1994).

<sup>26</sup>Sobel (1994), 798.

<sup>27</sup>Quine, “Two Dogmas of Empiricism” in Quine (1981), 42-6.

<sup>28</sup>I’d like to thank an anonymous reader for suggesting these comparisons.

ethics represents the application of coherentism to the justification of moral beliefs. It holds that one's moral belief  $p$  is justified insofar as  $p$  is part of a coherent system of beliefs, both moral and nonmoral, and  $p$ 's coherence at least partially explains why one holds  $p$ .<sup>29</sup>

Brink's coherentist moral epistemology is similar in many ways to my insistence that a person's genuine conception of the good should be coherent. However, PC makes no claims about the epistemic justification of a person's preferences. Objective theories of welfare might claim that the fact that A values  $p$  provides some reason for believing that  $p$  maintains preference-independent value, but PC itself is agnostic on this point and makes no claims about the evidential strength of a coherent system—evidential, that is, of anything beyond what a person genuinely prefers. It says only that A's coherent and complete beliefs about what is in A's interests constitutes A's genuine preferences and conception of the good. But this is as it should be. When we are interested in a person's genuine preferences, we are interested not in some external order of values to which our beliefs may or may not be reliable guides. Rather, we are interested only in *this person's* pro-attitudes: that which *she* believes is good for *her*. Of course, WBM claims that the fact that a person believes that a particular project  $p$  is good for her (along with the requisite comparative judgment about the life in which it is led) *renders* this project valuable enough to constitute the basic minimum. But this should be taken, simply, to be the result of an independently plausible claim about welfare (WBM), together with facts about what this person genuinely values.<sup>30</sup>

A second contrast should be drawn here. James Griffin, in *Value Judgment*, offers an account of prudential valuing that is in some ways similar to the approach I offer here.<sup>31</sup> On my view, a welfare-relevant preference is a belief about that which is good for a person, refined by means of coherence and complete testing. Griffin's view is similar. Griffin rejects the idea that "value judgments" are simply desires. According to Griffin, such judgments are a mix of belief and desire, a mix of "recognition and reaction", which includes a crucial cognitive element.<sup>32</sup> Furthermore, Griffin accepts a method

---

<sup>29</sup>Brink (1989), 103. Brink cites Rawls's reflective equilibrium procedure as embodying an essentially coherentist method of justification.

<sup>30</sup>This discussion gives rise to a number of questions about my view, including the semantic content of judgments about value ("if they're not about an external realm of value, what are they about?"). I shan't discuss these matters here, except to direct the interested reader to other venues in which I discuss them issues in much more detail. See Dorsey (2010b). For the purposes of this book, I leave aside such technical matters.

<sup>31</sup>Griffin (1996).

<sup>32</sup>Griffin (1996), 32-36.

of refining one's value judgments that seems to recall certain elements of my account of "value data".<sup>33</sup> For Griffin, the welfare value of something like "accomplishment" is known through a process of "recognition and reaction" of the importance of certain elements in human life; in this way, one might believe that a defect of a person's preferential profile is that this person has not engaged in the sort of reflection that would require confrontation with "value data" concerning a wide array of elements in human life.<sup>34</sup>

However, there is an important difference between Griffin's approach and my own that I should like to spend some brief time on.<sup>35</sup> Griffin writes: "For me to see anything as prudentially valuable, I must see it as enhancing life in a generally intelligible way, in a way that pertains to *human* life, not to any one particular person's life."<sup>36</sup> As noted above, my account of a person's genuine preferences denies this. On my view, A's genuine preferences are taken from A's beliefs about that which is good *for A*, not from A's beliefs about what is genuinely important to human life in any wider sense. Again, Griffin writes:

For anyone to see anything as valuable, from any point of view, requires being able to see it as worth wanting. This is a perfectly general requirement on values; it is the basis of the distinction between mere wanting and the sort of wanting that connects with values. One way to see something as worth wanting is to see it under the heading of some general human interest. . . . So long as one defines 'personal values' in terms of what a person wants or cares about, one retains too much of the taste [i.e., desire-based] model.<sup>37</sup>

Griffin's approach seems to have a novel solution to the problem of adaptive preferences: the overexhausted coolie, for instance, will generally not see

---

<sup>33</sup>See, especially, Griffin (1996), 25, 53-59.

<sup>34</sup>See Griffin (1996), 24-25.

<sup>35</sup>A few other differences should be obvious. First, I don't include the "reactional" element in a person's value judgments. Second, Griffin criticizes a model of ethical judgment that is simply coherentist. Griffin believes that such a model is of only slight benefit if we have no evidence that our ethical beliefs are generally reliable (Griffin (1996), 12-18). Of course, Griffin believes that the judgments we form about prudence and human interests are generally reliable (Griffin (1996), 66-67), and hence should be taken as good epistemic indicators of value. Again, I remain neutral on this point. I am interested only in the coherence and completeness of an individual's evaluative beliefs, which is necessary and sufficient for such beliefs to be representative of a person's genuine conception of the good.

<sup>36</sup>Griffin (1996), 27.

<sup>37</sup>Griffin (1996), 27-28.

overexhausted coolieism as “falling under the heading of some general human interest”.<sup>38</sup> Why, then, not embrace Griffin’s account, rather than my own approach, which focuses on A’s judgments about what is good *for A*?

Three considerations tell in favor of my view, at least for the purposes to which I put it here. First, we needn’t say that to see something as “worth wanting”, we must see it as valuable in terms of *general* human interests. Rather, one way in which I can see something as “worth wanting” is that I can *believe*, in a way that is supported by other such beliefs, that this thing *is good for me*. In this way, some things I want are not worth wanting: they are not things that I believe are good for me. All Griffin’s point requires is *some* extra-desiderative account of valuing against which we can test the evaluative quality of desires, not necessarily his particular account.

Second, I think there are very good reasons, in determining what a person’s genuine preferences are, to focus on what he believes is good for him, rather than what he believes is valuable for humans generally. For instance, we could imagine a person who believes that something is valuable for him (in a way that is coherent and complete), but who does not also see this as valuable for humans generally. A person engaged in the project of grass-counting may very well recognize that his pursuit is not for everyone, i.e., that it does not fall “under the heading of some general human interest”. However, it seems to me that this latter claim does nothing to diminish our judgment that the grass-counter *values* grass-counting. We may be tempted to say that grass-counting is not in anyone’s interest. But this is merely a call to restrict the relevance of this person’s genuine preferences in determining that which is in her interests, not to come to a different understanding of what she prefers or values. Furthermore, take a person who judges that a particular *p* *does* fall under the heading of a general human interest or need. But this person judges also (coherently and completely) that *p* is not good *for him*. Such a belief is not incoherent: this person could believe that he or she is relevantly different than human beings in a certain way, or may believe that it is good for him or her *to* be different in this way. I think we would say quite emphatically that this person does not value *p*.

Third, though we offer different interpretations of prudential judgments, it seems to me that our views are, ultimately, compatible. Griffin takes his account to be answering a very different question than the question my account of preferences addresses. Griffin holds that value judgments must be “reliable beliefs”, evidence of *objective* prudential values.<sup>39</sup> Griffin, like

---

<sup>38</sup>See, especially, Griffin (1996), 27.

<sup>39</sup>See Griffin (1996), 29-30, 53-59. Griffin himself objects to the “objective” terminology;

Brink, seeks to assess a more general moral epistemological question: what *sort* of beliefs are reliable indicators of an external realm of value? For Griffin, the “recognitionist” element of our value judgments are intended to be “a kind of sensitivity to something in the world”, which Griffin understands as those things that make human life—rather than any particular person’s life—better.<sup>40</sup> If that is correct, Griffin may very well believe that the best way to come to *true* judgments about welfare is to judge things against a background of general human interests, rather than with an eye toward what is good for me, in particular. PC, however, is not advanced as an epistemological guide to this external realm of values. Given this, PC and Griffin’s approach to value judgments need not be at loggerheads. As I have so far been at pains to argue, one plausible claim about objective prudential values is that the maintenance of a valued project is a central element of human well-being *in any life*.<sup>41</sup> One might assess this claim given its coherence with our beliefs about what is generally valuable for humans. But if this is correct, a further question arises: what, for A, is a valued project *for A*? The set of beliefs that results from the process embraced by PC helps to answer this latter question, but needn’t be construed as answering the wider question concerning whether a valued project is beneficial for persons, generally. Our methodologies respond to different concerns, and are appropriate to the concerns to which they respond.

### 3.4. Coherentism and Adaptation

PC avoids the problem of adaptive preferences by blocking the cognitive conditions that yield adaptive preferences. That which is shared by the cases of adaptive preferences we have so far considered is the extent to which facts about a given person’s actual life, including facts concerning what is available to that person influence the extent to which the person in question maintains a particular preference or pro-attitude. This fact explains any success the capabilities approach may have at avoiding adaptive preferences: we would expect the capabilities approach to correct preferences that arise as a result of a lack of basic capabilities or available styles of life. But when revising a person’s set of preferences according to PC, facts of availability

---

however, his view does conform to the idea of “objectivism” as I define it here. Griffin offers us a list of prudential goods, and holds that these goods are valuable “in any life” (Griffin (1996), 30). If so, Griffin has an objective view.

<sup>40</sup>Griffin (1996), 57-59.

<sup>41</sup>Griffin seems to accept a claim very much like this in identifying “accomplishment” as a central element of the objective list. See Griffin (1996), 29.

are rendered moot. Though it might be the case that Erin cannot live the life of a dancer, and hence develops an adaptive aversion to such a life, the *value data* that are used to revise her evaluative beliefs are not influenced by Erin's actual inability. Any relevant value datum used to revise her set of preferences includes a judgment of the comparative value of a particular life and its various goods given on the basis of a full experience of that life, not whether that life is available or is likely to occur given the facts about the world. Facts of actual availability will not influence the extent to which recalcitrant value data will cause revisions to her preferences. Because all lives are experienced, preferences that depend on facts of availability will be—other things equal—revised. For PC, the life that a person actually leads is simply one among many, and has no special status.

Furthermore, PC can also avoid problems of adaptation that are not simply due to facts of unavailability. Consider, for instance, examples of adaptation through “habit”, or perhaps certain forms of social indoctrination. *If* these preferences are not expressive of an individual's true conception of the good, we should expect that after having undergone complete testing against all metaphysically possible value data, one's preference for  $p$  or  $q$  as a result of social pressure, indoctrination, or habit will be revised as recalcitrant. When experiencing value data, one experiences lives in which one's habits are markedly changed, one's social pressures are very different, and in which one is not indoctrinated. *If* these preferences are genuinely adaptive, we should expect that once these habits, pressures, etc., are removed, the person in question will no longer prefer  $p$  to  $q$ —or, at the very least, will prefer  $p$  to  $q$  in a way that is reflective of her genuine conception of the good.

Of course, adaptive preferences are an ineradicable feature of value data: the *experience* of living a life will surely present many *examples* of biased or adaptive preferences. But the preferences one has during any particular life are not necessarily representative of one's conception of the good, properly conceived. Though, in life  $l$ , Erin might strongly prefer painting to dancing, Erin's coherent and complete conception of the good can declare this preference irrelevant to her welfare. And it might do so by comparing the assessment of  $l$  against an assessment of life  $k$ , in which a life of dance is available and experienced. The relevant value datum here is likely to declare that dancing is to be preferred, sufficient to override a judgment from within  $l$  that a life of painting is to be preferred.<sup>42</sup>

---

<sup>42</sup>However, were Erin to judge her life as a dancer worse than her life as painter from the perspective of her coherent and complete conception of the good, being a painter *is*

It might be objected that PC is still beholden to the (perhaps adaptive) preferences of a person's original set of evaluative beliefs in one sense: conceptions of the good are revised conservatively, i.e., by minimal mutilation, which requires that we override a person's most strongly held judgments only as a last resort. It is not at all guaranteed that preferences that are originally a result of facts of availability will not be among the person's strongest held beliefs from the perspective of her conception of the good. After all, over the course of time and psychological change a person might come to strongly identify with preferences that were originally cases of adaptation, or other forms of psychological distortion. If so, confrontation with value data is insufficient to remove biased preferences because this reflection will surely be done in light of these very preferences. However, PC has resources to alleviate this worry. After all, *for a preference to be adaptive*, there must be some reason for thinking that this preference arises merely given facts about what is or is not available or other various facts about the particular life a person leads (including habit, social pressures, indoctrination, etc.). Because Erin's biased preferences arise as a result of, and are supported by, these facts, Erin will prefer achievement in dance when these facts are rendered moot. In experiencing value data, there is no longer any feature of the world or of the life one actually lives that would allow biased preferences to develop, or recommend preference revision. All beliefs are subject to revision, and given the nature of adaptive preferences, they are likely to be revised in light of complete testing against value data, insofar as a person's *particular life* is simply one among many metaphysically possible lives.

Thus, returning to WBM, PC solves the problem of adaptation as follows: any judgment that a particular life is worth living, or that a particular project contributes to the worth of a particular life, will not be made on the basis of facts of unavailability habit, indoctrination, etc. Such judgments, if they are to reflect an individual's genuine conception of the good, must survive the process described here. Thus we should not fear anyone's maintenance of the basic minimum on the basis of adaptive preferences. No preferences, pro-attitudes, or other evaluative judgments that are held simply on the basis of facts of the particular facts of a person's actual life will determine whether a given person maintains the basic minimum.

### *3.5. Too Low, Part Two: Deep Adaptation and Preference for the Worse*

---

a better life for Erin. Furthermore, if the respective value data judge these lives to be equivalent in value, Erin's conception of the good will also reflect this fact. This is the correct answer.

PC makes as much headway on the problem of adaptive preferences as Nussbaum’s capabilities approach. Recall that Nussbaum’s approach treats adaptation as a result of facts of unavailability or narrow sets of capabilities. In correcting these facts, Nussbaum avoids the problematic suggestion that people who possess the basic minimum live shabby lives as a result of adaptive preferences—for Nussbaum, the provision of capabilities is mechanism enough to overcome adaptation. WBM can do just as well. PC, and hence WBM, holds that preferences held simply on the basis of facts about a given person’s actual life (such as availability, habit, indoctrination, etc.) will not influence that which constitutes, for them, the basic minimum. Hence, for WBM, people who maintain adaptive preferences will not maintain the basic minimum simply on the basis of fulfilling these adaptive preferences. But there is an important dimension of the problem of adaptive preferences that has so far been left out: adaptation (through whatever mechanism) can give rise to preferences that are *central* elements of a person’s conception of the good. So far I have been assuming that adaptive preferences are preferences that do not express an individual’s true values. But what if habit, indoctrination, or simple facts of unavailability—perhaps as a result of the sands of time—come to be ingrained as a central part of an individual’s conception of the good, which complete testing will not revise?

This question illustrates a difference between two forms of adaptive preferences (to which I briefly alluded in the introduction to this chapter). Call “shallow adaptive preferences” those preferences that are adaptive insofar as they do not reflect an individual’s true conception of the good. Call “deep adaptive preferences” preferences that, as a result of various mechanisms of adaptation, are not simply a product of facts of unavailability, but are rather a *central* element in a given person’s conception of the good, which the processes central to PC will not alter. Given that PC relies on revision of a person’s conception of the good using a rubric of minimal mutilation, there is no guarantee that previously adaptive preferences will not themselves be the most central, important bits of a conception of the good, ruling other evaluative judgments simply recalcitrant. Though PC can solve the problem of shallow adaptive preferences, it does nothing to solve the problem of *deep* adaptation.

In evaluating WBM in light of deep adaptation, however, one must always keep in mind the extent to which its competition can do better. In considering the possibility of deep adaptation, or preferences that are “adaptive through and through”, Nussbaum effectively denies their existence. According to Nussbaum:

the human personality has a structure that is at least to some extent independent of culture, powerfully though culture shapes it at every stage. As the Greek philosophy Sextus Empiricus wrote, “In the person burdened by hunger and thirst, it is impossible to produce by argument the conviction that he is not so burdened.” Desires for food, for mobility, for security, for health, for the use of reason—these seem to be relatively permanent features of our makeup as humans, which culture can blunt, but cannot altogether remove.<sup>43</sup>

Of course, we may deny that Nussbaum’s psychological claims are correct. If so, deep adaptive preferences are no less a problem for her approach than for mine. But if we accept this psychological claim—which itself seems plausible—this helps to alleviate deep adaptation for welfarism, as well. Few people “burdened by hunger and thirst”, or who live lives of subjugation, destitution, etc., etc., will regard any project lived in such a life as valuable: few people would regard these lives as worth living or of positive rather than negative value, after having undergone complete confrontation with all possible value data.

Furthermore, the problem of deep adaptation *in and of itself* is neither a problem for Nussbaum’s view, nor for mine. To see this, consider two cases. First, imagine that Erin’s preferences shift from  $t_1$  to  $t_2$  as a result of successful preference engineering. Now, at  $t_n$ , Erin’s adaptive preference for painting has become a central, ingrained aspect of her conception of the good, so ingrained that complete testing against all metaphysically possible value data will not reverse her preference for painting over dancing. In this case, it seems to me difficult to declare that dancing would be better for Erin than painting. After all, she strongly prefers it, and though this preference did not express her genuine conception of the good at  $t_2$ , *it does at  $t_n$* . Hence, it seems to me, there is little reason to believe that Erin’s *deep* adaptive preferences are irrelevant to her welfare in the same way that Erin’s  $t_2$  *shallow* adaptive preferences are. Once Erin’s preferences shift in a way that genuinely reflects her values, there is little call to index her welfare to what her preferences *used to be*. Her current preferences are genuine. However, now compare Erin to Celie. One might imagine that Celie, through a lack of self-worth, comes to value her shabby life. Assume also that this adaptive preference is *deep*: it is a central part of her conception of the good. In this case, deep adaptation is a problem: Celie’s preferences, or so it would seem, are not a good guide to that which is in her best interests.

---

<sup>43</sup>Nussbaum (2000), 155.

What distinguishes Erin and Celie? It is not deep adaptation: both maintain preferences that are deeply adaptive. Rather, deep adaptation seems problematic only when these preferences are also an example of *preference for the worse*. As I stated before, one prefers the worse insofar as one prefers that which is objectively worse to that which is objectively better. Erin clearly does not prefer the worse; this fact seems to explain our reaction to her deep adaptive preferences. Celie, on the other hand, does.

If this is correct, deep adaptation is not *in itself* a problem for WBM. But Celie's case shows that even though deep adaptation is not a *per se* problem, preference for the worse *is*. There are certain projects we may regard as being "beneath" the basic minimum simply because they are independently of shabby quality: they are objectively bad rather than objectively good. One could imagine that "being a dominated housewife" (like Celie) might very well be a global project that Celie genuinely values. And if that's right, WBM seems to implausibly suggest that Celie maintains the basic minimum. One might be tempted to saddle Nussbaum's approach with exactly the same problematic conclusion. After all, if Celie values her shabby life in a *deeply* adaptive way, wouldn't she continue to live such a life even with greater capabilities? Generally speaking, I think the answer is yes. But WBM has a problem that Nussbaum's view escapes: we might imagine that Celie *prefers* to be non-dominated, and given the capability for non-domination, would avoid it (though being a dominated housewife remains a valued project). If this is correct, the capabilities approach seems to entail that Celie, in maintaining the basic minimum, will avoid being a dominated housewife. WBM, on the other hand, seems to include no such mechanism. In living a life as a dominated housewife, Celie will maintain the basic minimum though she may prefer the non-dominated life. Looked at in this light, the problem of preference for the worse becomes a new way of stating the objection I discussed last chapter: that WBM identifies a basic minimum that is simply *too low*.

However, this objection can be rebutted. Notice that either theories of welfare will, or will not, accept the possibility of "content-based assessment" of valued projects. In other words, some will believe that, in terms of their welfare value, valued projects can be *assessed by the objective value of the projects themselves*. On this sort of view, some projects will maintain preference-independent value, others will not, and preference-independent value is a *pro tanto* determinant of the intrinsic value of such projects for persons. Other theories of welfare will refuse to evaluate valued projects based on content, but will instead rely simply on the preferences of the individual whose projects they are.

The dispute between welfarists who accept content-based assessment of valued projects and those that do not is long, deep, and something I won't settle here. However, it is certainly true that *either* valued projects can be assessed by their content or they cannot. But in either case, a welfarist basic minimum survives.

### 3.5.1. Preference for the Worse and Content-Based Assessment

Some will accept that valued projects can be assessed for their welfare value on the basis of content. Some valued projects (like, e.g., being an excellent dancer) will maintain preference-independent value, others (like, e.g., being a dominated housewife) will not. However, if we accept this sort of view, an obvious and immediate response to the problem of preference to the worse arises. Recall that an essential element of any basic minimum is the ability to draw a “dividing line” between lives that maintain some morally or evaluatively essential feature and lives that do not. WBM insists that the dividing line is between lives that do, and lives that do not, maintain a valued project. But if one accepts the possibility of content-based assessment, one might draw the dividing line differently: between lives that maintain valued projects of *preference-independent value*, and those that do not. Consider, then:

*Welfarist Basic Minimum 2* (WBM2): A achieves the basic minimum over the course of a life  $l$  if and only if A successfully achieves or maintains a valued project of sufficient preference-independent welfare value in  $l$ .<sup>44</sup>

WBM2 insists that valued projects must be of preference-independent value if their fulfillment is to constitute the basic minimum. WBM2 solves the problem of preference for the worse. One might declare that Celie, no matter how deeply she values a life in which she is dominated or treated shabbily, cannot maintain the basic minimum simply given the worthlessness of her project. Of course, WBM2 insists that to achieve the basic minimum Celie must successfully maintain a valued project. But a particular valued project *cannot* be of welfare value sufficient to maintain the basic minimum unless that project is also of sufficient preference-independent welfare value. This solves the problem of preference for the worse.

In interpreting WBM2, two points are worth noting. First, the account of preference-independent value is left open. One might insist that preference-

---

<sup>44</sup>If one prefers the time-relative version of WBM, WBM2 can be translated *mutatis mutandis*.

independent value is to be identified as being accordance with human dignity, nature, or “truly human functioning” (*modulo* my arguments against such possibilities in the previous chapter). Alternatively, one might identify the preference-independent value of valued projects by means of some form of objective list. Second, WBM2 is ecumenical between substantive theories of welfare, just as WBM was. For instance, Weak Strong Subjectivism holds that though the fact that  $p$  is valued by A is necessary and sufficient for  $p$  to be of value for A, Weak Strong Subjectivism can declare that the basic minimum requires one not simply to achieve a project one values, but also requires a valued project of preference-independent value. Celie, in other words, is benefited by being a dominated housewife because she values this project. But without maintaining a project that is of sufficient objective value, her life—no matter how many other things she achieves—is not good enough to maintain the basic minimum. Weak Subjectivism, as opposed to Weak Strong Subjectivism, might accept an even stronger claim. Weak Subjectivism needn’t declare that Celie’s achievement of the projects she actually values is of any benefit to her at all: Weak Subjectivism could say that Celie is benefitted *only* when obtaining a valued project of preference-independent value. Both Weak Subjectivism and Weak Strong Subjectivism can accept WBM2, and hence can plausibly respond to the problem of preference for the worse. Furthermore—as must be fairly obvious—an objectivist can also accept WBM2. Indeed, an objectivist can either agree with Weak Strong Subjectivism—and claim that being a dominated housewife does benefit Celie, but not enough to obtain the basic minimum—or agree with Weak Subjectivism—and claim that the fulfillment of Celie’s preferences for the worse are of no independent welfare value. WBM2 remains widely ecumenical.

WBM2 is thus an option for theories of welfare that accept content-based assessment of valued projects. I leave this possibility fully open, and will indicate in the remainder of the book points at which further discussions will require modification in light of WBM2 rather than WBM. However, I do want to offer a few words in favor of accepting WBM even for theories that allow content-based assessment of valued projects. For starters, whatever else her life is actually like, Celie’s life has a self-directed meaning that she regards as being “worth it” rather than not, good rather than bad, etc., in a way that reflects her true conception of the good. As I have so far argued, this is no minor welfare achievement. There may be other very good reasons to assist Celie, even to regard her situation as morally urgent. But she has something important that many other people lack: a life that maintains a self-directed meaning, one that, given her genuine values, contributes to

her own assessment of the worthiness of her life. However her values come about, whatever her values really are, they are deeply and essentially *hers*. Thus the difference between deep adaptation and shallow adaptation is of the essence. For people who lack shallow adaptive preferences any life they lead that maintains the basic minimum is genuinely valued. The activities of this life can be said to have a shared meaning, one that this person believes is valuable for her, lived within the context of a life she (coherently and completely) believes is worth living. Viewed in such a light, I have a very hard time accepting the claim that WBM sets the basic minimum too low.

Fair enough, it might be said, when viewed in isolation. But is this really plausible for Celie, even if she regards her project as contributing to an assessment of her life as worth living? I think so. Consider again a simple trade-off of the sort we considered in the previous chapter. Take Glen, who has no valued project at all, but could potentially maintain such a project. Now compare Glen to Celie, to whom we could grant a further valued project of preference-independent value. It seems to me that if we are concerned only with the *promotion of valuable states*, distributional priority ought to go to Glen. Celie, rather than Glen, maintains a life of self-directed meaning that she actually regards as valuable. Glen does not. It seems to me that the dividing line between people who do and do not maintain valued projects is of greater moral significance than the dividing line between people who do and do not maintain projects of preference-independent value. If that is correct, we should accept the claim that Celie maintains the basic minimum. Again, the conceptual structure of any basic minimum looms large: the basic minimum *ought* to be low, low enough to plausibly support upward distribution.

One final problem should be explored. In the face of the problem of preference for the worse, WBM seems to have a frightful moral implication. It would appear that the moral reason to establish the basic minimum could morally justify treating a person like Celie *so shabbily*, or justify subjecting her to such strong brainwashing, that the shabby projects she actually ends up maintaining are a *central* (or “deep”) part of her conception of the good, rendering her activities constitutive of the basic minimum. But it seems wrong to say that such actions could be morally justified. In response, WBM can sensibly reject the troubling claim that a person might be morally permitted to engage in such brainwashing for the sake of the achievement of the basic minimum. If we accept the possibility of content-based assessment, we can quite sensibly claim that Celie would be better-off maintaining the basic minimum with a project of preference-independent value than she would be maintaining the basic minimum *as a dominated housewife*. Insofar

as one always has a moral reason to provide a person with a better, rather than a worse, life, one always has greater reason (given that we are assuming the possibility of content-based assessment of valued projects) to provide an individual the achievement of an objectively valuable valued project rather than one that is objectively valueless. If that is right, even if we accept WBM, there is far stronger reason simply in terms of the promotion of valuable states to grant a *more* valuable achievement of the basic minimum than a less valuable achievement of the basic minimum. In addition, we might perfectly well accept reasons not to brainwash or indoctrinate people into living lives they would not otherwise have valued. One might suggest, for instance, that tinkering with someone's conception of the good in this way is a fundamental lack of respect or a violation of that person's autonomy.<sup>45</sup> We can embrace sensible moral restrictions on brainwashing and shabby treatment even if such brainwashing is a means to the achievement of the basic minimum. Again, the promotion of valuable states need not be all there is to morality.

These considerations seem to me to provide good reason for any substantive theory of welfare, even those that accept content-based assessment, to accept WBM rather than WBM2. But I cannot pretend to have offered anything like a knock-down argument here. Hence I leave WBM2 open for those who are genuinely concerned about preference for the worse. Very little in the remaining chapters will require substantive translation. Whether one accepts WBM or WBM2, however, is less important than the general lesson: a welfarist basic minimum can survive the challenge of preference for the worse, or the claim that the basic minimum is set too low.

### *3.5.2. Preference for the Worse and No Content-Based Assessment*

Now assume that we reject content-based assessment. If so, the only relevant facts concerning the merits of valued projects are the extent to which they are genuinely valued. And if *that's* right, Erin's project and Celie's project are of equivalent welfare value (assuming that they are valued to the same extent). Under these conditions, preference for the worse is nonsensical: no valued projects are worse than any others. If such a view is correct, WBM seems the obvious choice.

One might think this is a problem. After all, this would entail that Celie maintains the basic minimum no less than Erin. Surely this is wrong! But if we are genuinely rejecting the claim that Erin's project is any better than

---

<sup>45</sup>See, for instance, Pogge (2001), 37.

Celie's in terms of its content, this is exactly the *right* conclusion. Few would disagree that Erin's great achievement in painting would constitute a life that maintains the basic minimum. But to say that a life that is *just as good as Erin's*, i.e., Celie's, does not maintain the basic minimum one must say that Erin maintains, but Celie does not, some particular non-welfare good  $g$  that is essential for the basic minimum, but that makes *no difference whatsoever* to the quality of Erin's life. This smacks of the sort of fetishism displayed by the capabilities approach I discussed in Chapter One: a commitment to the provision of non-welfare goods even when such provision will do nothing to alter the relative quality of lives. But this is surely morally problematic if anything is. If we really agree that there can be no content-based prudential assessment of projects, so long as they genuinely value their projects to an equivalent degree, Erin and Celie live lives that are equally good, and hence it seems correct to say that they equally maintain the basic minimum. WBM survives.

That Erin and Celie live equally good lives, simply on the basis of their own conceptions of the good, may sound absurd, even obscene. This, it seems to me, is less a problem for WBM than it is for the claim that we should reject content-based assessment of welfare goods. However, as I argued in the previous section, there remains good reason to accept WBM even if we accept the possibility of content-based assessment. WBM guarantees not just a life that a person regards as worth living, but a life that includes a valued project that unifies, and provides a self-directed meaning to, one's life and activities. This seems to me a significant welfare achievement, one that we would be unwise to allow the basic minimum to exceed. I leave WBM2 open for those who disagree.

### 3.6. Conclusion

Recall the welfarist's dilemma: either one accepts *The Basic Minimum as Proto-autonomy*, and one succumbs to the problem of adaptive preferences, or one rejects *The Basic Minimum as Proto-autonomy*, and one commits to forcing people to live lives they do not value. But I have shown that this is in fact no dilemma whatsoever. In fact, one can respect individuals' conceptions of the good only if one *also* avoids adaptive preferences. And we respect conceptions of the good and avoid adaptive preferences by better understanding the nature of conceptions of the good. To this end, I have offered a theory of a person's true conception of the good: PC. In addition, I have shown that there is very little reason for a welfarist theory of the basic minimum to fear the problem of preference for the worse. First, pref-

erence for the worse is not a genuine problem for WBM, even if we allow for content-based assessment. Second, even *if* we accept that preference for the worse is a problem for WBM, it is not a problem for welfarist theories of the basic minimum on the whole: one can simply accept WBM2. No matter the content of an individual's genuine preferences, maintaining a valued project is a crucial and central welfare achievement; it is the maintenance of a life of unified, self-directed meaning one *genuinely* values.